AN ADAPTIVE APPROACH TOWARDS CONTENT-BASED IMAGE RETRIEVAL

Jana Urban, Joemon M. Jose and C.J. van Rijsbergen

Department of Computing Science, University of Glasgow, Glasgow G12 8QQ, UK
{jana,jj,keith}@dcs.gla.ac.uk

ABSTRACT

We propose and evaluate an adaptive approach towards content-based image retrieval (CBIR), which is based on the Ostensive Model of developing information needs. We use ostensive relevance to capture the user's current need and tailor the retrieval accordingly. Our approach supports content-assisted browsing, by incorporating an adaptive query learning scheme based on implicit user feedback.

Textual and colour features are employed to characterise images, which are combined using the Dempster-Shafer theory of evidence combination. Results from a user-centred, task-oriented evaluation show that the ostensive interface is preferred over a traditional interface with manual query facilities. Its strengths lie in its ability to adapt to the user's need, and its very intuitive and fluid way of operation.

1. INTRODUCTION

The *semantic gap* has become a buzzword in Content-based Image Retrieval (CBIR) research. It refers to the gap between low-level image features and high-level semantic concepts. Since the low-level features do not directly reflect the user's high-level perception of the image content, the query formulation is even more difficult than in text IR systems. In addition, the underlying search need is dynamic and evolving in the course of a search session. Today's CBIR systems fail to deal with the dynamic nature of search needs.

In this work, we propose and evaluate an adaptive image retrieval technique, which is based on the *Ostensive Model (OM) of developing information needs* [3]. The OM recognises and addresses the issue of *dynamic nature of information needs*, and has the advantage of allowing for an intuitive and user-centred search process.

The remainder of the paper is organised as follows. Section 2 provides an account of the motivations for our approach. It describes current relevance feedback techniques, and compares those to the notion of ostensive relevance. The systems we used for the evaluation are outlined in section 3 and section 4 describes the proposed adaptive query learning scheme. The experimental methodology is detailed in section 5, followed by a review of our experimental results in section 6. Finally, section 7 concludes the paper.

2. BACKGROUND / MOTIVATION

One of the major issues in information searching is the problems associated with initiating a good query. However, it has been identified that searchers find it hard to generate a query due to the following reasons. Firstly, searchers do not know how the documents are represented and even if they do, they may find it hard to formulate the query in terms of the low-level image features used in many systems. Secondly, most often, the underlying information need itself is typically vague (*"I don't know what I'm looking for, but I'll know when I find it"* [14]). As a result, the search process is explorative in nature, in which a user's need may change dramatically due to the exposure to new information. This often leads the user to reformulate the initial query to either make it more precise after having gained some knowledge about the collection make-up, or steer it in different directions after having seen other interesting documents, or a combination of both.

2.1. Relevance Feedback

In order to alleviate the query formulation problem, a popular approach is to incorporate relevance feedback in the retrieval system. The idea of incorporating relevance feedback first emerged in text retrieval systems (e.g. [10]), and has been studied since. It is even more valuable in the image domain: a user can tell instantaneously whether an image is relevant with respect to the current context, while it takes substantially more time to read through a text document to estimate its relevance.

It is regarded as an invaluable tool to improve CBIR systems, for several reasons. Apart from providing a way to embrace the individuality of users, they are indispensable to overcome the *semantic gap*. Rather than trying to find better techniques and more enhanced image features in order to improve the performance of what has been referred to as *"computer-centric"* systems [11], it is more satisfactory to the user to exploit human computer interaction to refine high level queries to representations based on low-level features. A comprehensive study of existing relevance feedback techniques in image retrieval can be found in [15].

Different methods have been adopted on the basis of often diverging assumptions. One major variance is *how* the information about the user's judgment of relevance is gained. One can distinguish two distinct approaches: *explicit* and *implicit* relevance feedback. Explicit relevance feedback asks the user to explicitly state whether a returned document is relevant or not. This additional task is often considered as a cognitive burden to the user, since it is difficult for most users to assess the degree of relevance of one document in terms of a numeric value, which presumes considerable knowledge of the retrieval environment. This problem parallels with the query formulation problem mentioned before. For this reason, a less-distracting possibility to gain relevance feedback is implicitly from the users, simply by observing their interaction with the system.

Another assumption underlying nearly all current relevance feedback techniques is that a user's information need is static and there is no provision for updating user's judgements. However, this is a rather simplifying view of the real-world. Not only are the user's *actions* time-dependent—resulting in giving inconsistent feedback, but even more importantly, the user's *goals* are also time-dependent and might change either gradually or quite abruptly. The trigger for such changes is most often a result of having come across something interesting that they have not even considered at the beginning of the search.

For all those reasons, the proposed approach is based on the *Ostensive Model*, which deals with the problems discussed above. It captures *"the intentionality of an information need that is assumed to be developing during the searching session"* [2]. While most current systems ask for explicit relevance feedback (e.g. [11, 8, 9]) the relevance judgements in this model are obtained implicitly, by interpreting a user's selection of one document over others as an indication that this document is more relevant. Instead of using this feedback for feature selection (e.g. [8]) or feature relevance weighting (e.g. [11]), the query itself is learnt and subject to adaptation. On the basis of the selected documents, the system creates a new query consisting of a combination of those documents' features. This query adapts in every iteration of the retrieval process. The details of the model will be described in the following section.

2.2. Ostensive Relevance

The Ostensive Model (OM) of developing information needs was initially proposed by Campbell and van Rijsbergen [4]. It combines the two complementary approaches to information seeking: query-based and browse-based. It supports a query-less interface, in which the user's indication of the relevance of an object—through pointing at an object—is interpreted as evidence for it being relevant to his current information need. Therefore, it allows direct searching without the need of formally describing the information need.

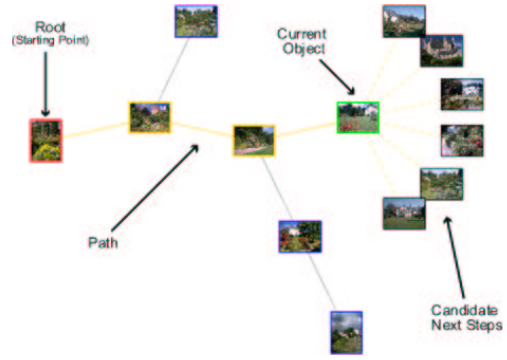


Fig. 1. The ostensive path

This reflects another important issue, not previously addressed in traditional CBIR systems: the *dynamic nature of information needs*. Through exposure to new objects, the user's context is changing and their knowledge state is developing. By accepting that the user's need is dynamically changing during a search session, the notion of relevance also has to be adapted. Hence, the OM adds a temporal dimension to this notion. A recently selected object is regarded more indicative to the current information need than a previously selected one. So, in this sense, the degree to which a document is considered relevant is continuously updated to reflect the changing context. Campbell's definition of Ostensive Relevance summarises the main points [3]: *"The **Ostensive Relevance** of an information object is the degree to which evidence from the object is representative/indicative of the current information need."*

An ostensive browser facilitates a new form of interaction [3, 4]. A user starts with one example document as the query, and as a result a new set of candidate documents (top ranking documents according to the similarity measure used) is presented to the user. As a next step, the user—through selecting one of the returned documents—updates the query, which now consists of the original document and the selected document of the set of returned candidates. After a couple of iterations, the query is based on a path of documents. Similarly to the Path Model described in [5] for activity-centred information access, emphasis is set on the user's activity and the context, rather than the predefined internal representation of the data. A path represents the user's motion through information, and taken as a whole is used to build up a representation of the instantaneous information need.

Since the whole path is visible to the users, they can jump back to a previous object along the path if they get the feeling that they are stuck or moving in the wrong direction. From there a new path can be explored, starting from the original object (the root) and the newly selected object. The resulting paths form a tree-like structure, originating from one root and branching at various objects (see Fig. 1).

The OM thus captures the user's developing information need during a search process, and incorporates the uncertainty, which necessarily exists due to the imprecise awareness of one's own need and the difficulties of expressing it.

No effectiveness evaluation of the ostensive approach has ever been undertaken. In addition, in [4, 3, 2] the OM was implemented with the Binary Probabilistic Model using textual features only. There was no provision for image features, such as colour or texture, for retrieval purposes. In this work, we extend the OM with an adaptive learning scheme based on multiple features. In addition, the effectiveness of this approach is evaluated in a user-centred setting. In the next section, we describe the different systems employed for the evaluation.

3. THE SYSTEMS

The systems use two distinct features: *text* and *visual*. The text feature is extracted from the keyword annotations of the images, and the visual feature is based on global colour histograms. An image is represented by a two-dimensional feature vector, which is a term vector (text feature) and a histogram bin vector (colour feature) respectively. The term vector is weighted by the $tf \times idf$ weighting scheme. The similarity between documents is determined by calculating the two similarity values for each feature. In the case of text similarity, the *cosine measure* [12] is used whereas the visual similarity is determined by *histogram intersection* [13].

3.1. The Interfaces

3.1.1. The Ostensive Browsers

Two versions of the ostensive browsing approach have been implemented: One with a pure ostensive browsing scheme (Fig. 2(c)) and one allowing explicit feedback with ostensive browsing (Fig. 2(b)). In both systems the user starts with an image in the browse panel (in Fig. 2(b)-2). The initial image is obtained in a pre-keyword search from which the user is given the opportunity to choose an image to explore further. When selecting an image, the system returns a set of most similar images as candidate images. We chose to present six images as new candidates. Of those candidates, the user clicks on the most appropriate one. At this stage, the system computes a new set of similar images based on an adapted query and present it to the user. As in figure 2(b) & (c), this process creates a path of images, which is represented in the interface. At any point the user can go back to previously selected images in the path and also branch off, by selecting a different candidate. Since the screen space is very limited the different paths are often overlapped resulting in a large degree of clutter, a fish-eye view as an alternative (see Fig 2(c)) is also provided. The user can alter between these views during the search.

To view details of the image, there is the possibility of viewing a single selected image in full-size in a separate panel (in Fig. 2(b)-3). It also contains some meta-data about the document, such as the photographer, title, date, and description. In between the full-size view and the thumbnails, a quick view is shown as a popup when the user moves the mouse over a thumbnail in the browse panel.

Both browsers attempt to adapt the query based on the user's implicit feedback. We provided two slightly different versions of the Ostensive Browser to allow for different levels of control. The **Pure Ostensive Browser** (POB) (Fig. 2(c)) does not allow for any control of feature terms or weighting between the features. The system automatically adapts the query and also the feature weights.

The interface for the **Controlled Ostensive Browser** (COB) also provides options for selecting the features and their associated weights (in Fig. 2(b)-1). It provides feedback about the search terms the system used to obtain the currently shown candidates. The automatically selected terms (the strategy of the selection is described in section 4), can be changed by the user and thus the current candidates are exchanged for the ones resulting from the updated query. Another aspect of control is the adjustment of the feature weights. The user can control the weights between the two features by means of a slider.

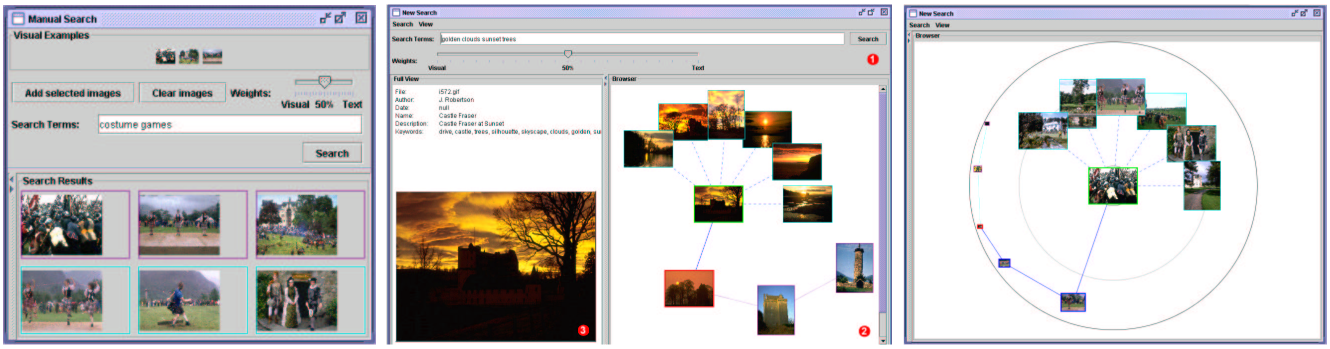
3.1.2. The Manual Query System

As a baseline system, we used a **Manual Query System** (MQS) (Fig 2(a)) resembling a 'traditional' image retrieval system, which returns a set of relevant images in response to a user-given query. A query can be formulated by a set of keywords and/or one or more images as 'visual examples'. The user can also set the weighting between the two features. If the user is not satisfied with the results returned by the system, he has to alter his query and so forth.

4. QUERY ADAPTATION TECHNIQUES

In the course of a search session, a user creates and moves along a path of images. In every iteration, the path changes and the query needs to be adapted accordingly. The objects are treated as evidence of the user's information need, with a changing degree of uncertainty associated to each object: the older the evidence, the more uncertain it is that it is still indicative of the *current* information need. The degree of uncertainty is represented by the ostensive relevance profile.

The query is comprised of the weighted sum of evidence found in the path objects. The evidence as far as the retrieval system is concerned is collected from the documents' features. Therefore, we propose an adaptive query learning scheme, in which the images on the path contribute to the new query. Each contribution is weighted with respect to an



(a) MQS (b) COB (c) POB

Fig. 2. The interfaces.

ostensive relevance profile [2](computed by a function $\frac{1}{2^k}$, with k meaning the position in the path, starting at the most recently selected object), which decreases with increasing distance to the current object. This is done for each feature separately.

Text Query: For the text feature, one can create a new query vector, by updating the term's intrinsic weights (e.g. idf) with the ostensive relevance weights resulting from the ostensive profile. The query vector then consists of the union of the set of terms, which appear in any of the documents on the path. The term's original weight is multiplied by the sum of ostensive relevance values for all documents in which the term appears:

$$w_t = idf_t \times \sum_{\substack{i=1 \\ t \,\in\, D_i}}^{l_p} (ORel_i \;\times\; tf_t(D_i)) \tag{1}$$

where w_t is the resulting weight of term t in the query vector, idf_t the term's idf value, l_p the length of the path, D_i the document at position i in the path, $tf_t(D_i)$ the term frequency of term t in document D_i, and $ORel_i$ the ostensive relevance weight at position i. The ostensive relevance weights are normalised, thus $\sum_{i=1}^{l_p} ORel_i = 1$.

Hence, the terms from the images are ranked with respect to the relevance profile and their corresponding *idf* values. A new query vector is computed based on the top four terms.

Histogram Query: A straight-forward approach in accordance with other query-point movement techniques (e.g. [9]) for constructing the query histogram is a linear combination of the constituent histograms and the ostensive relevance weights:

$$H_Q = \sum_{i=1}^{l_p} (ORel_i \times H_{D_i}) \tag{2}$$

The resulting query histogram H_Q is comprised of the bins computed as the weighted sum of the path documents' bins. It can be interpreted as the weighted 'centroid' of the corresponding histograms.

4.1. Final Evidence

Two queries representing each feature are issued to the system, returning two result lists with different scores based on the respective similarity measure for each feature. A means to combine the results to obtain one single ranked list is the *Dempster-Shafer Theory of Evidence Combination*.

The Dempster-Shafer mechanism has been widely used in the context of IR to combine information from multiple sources [6]. The advantage of Dempster's combination rule, is that it integrates degrees of uncertainty or trust values for different sources. For two features Dempster-Shafer's formula is given by:

$$\begin{aligned} m(\{d_i\}) \;&=\; m_1(\{d_i\}) \times m_2(\{d_i\}) + \\ &\quad\; m_1(\Theta) \times m_2(\{d_i\}) + m_1(\{d_i\}) \times m_2(\Theta) \end{aligned} \tag{3}$$

and

$$m(\Theta) = m_1(\Theta) \times m_2(\Theta) \tag{4}$$

where $m_k(\{d_i\})$ (for $k = 1, 2$) can be interpreted as the probability that document d_i is relevant with respect to source k. The two sources in our case correspond to the similarity values computed from the text and colour feature respectively. Θ denotes the global set of documents, and $m_k(\Theta)$ represents the uncertainty in those sources of evidence (also referred to as un-trust coefficients):

$$m_k(\Theta) = 1 - strength_k \tag{5}$$

where:

$$strength_k = \frac{\sum_{i=1}^{l_p} m_k(\{d_i\})}{\sum_{i=1}^{l_p} m_1(\{d_i\}) + \sum_{i=1}^{l_p} m_2(\{d_i\})} \tag{6}$$

$strength_k$ corresponds to the trust in a source of evidence k. That is, how much the given source represents the information need. The piece of evidence in this case is the calculated similarity values for the two features $m_1(\{d_i\})$ and $m_2(\{d_i\})$. The resulting $m(\{d_i\})$ is thus the combined belief for document d_i. Formulae 3&4 are a simplified version of Dempster-Shafer theory for IR purposes. Furthermore, it can easily be extended to accommodate more than two sources.

5. EXPERIMENTAL METHODOLOGY

It has been argued that traditional IR evaluation techniques based on precision-recall measures are not suitable for evaluating adaptive systems [7]. Thus in order to evaluate the systems, we used a task-oriented, user-centred approach.

In our evaluative study, we adopted a randomised within-subjects design, in which 18 searchers used three systems on three tasks. The independent variable was system type; three sets of values of a variety of dependent variables indicative of acceptability or user satisfaction were to be determined through the administration of questionnaires.

To reduce the effect of learning from one system to the other, the order of the systems and tasks was rotated according to a Greco-Latin square design. For the purpose of the experiment we employed the photographic collection, containing 800 photographs, created from the photographic archive of the National Trust for Scotland [6].

5.1. Tasks & Systems

In order to place our participants in a real work task scenario, we used simulated work task situation as conducted in [7]. This scenario allows the users to evolve their information needs, in just the same dynamic manner as such needs might be observed to do so in participants' real working lives. There were three tasks and each task involved at least two searches. The tasks were chosen to be of very similar nature, in order to minimise bias in the performance across the systems.

The Ostensive Browsers (sec. 3.1.1) were evaluated against the 'traditional' image retrieval system MQS (sec. 3.1.2), which supports manual query facilities.

5.2. Hypothesis

Our experimental hypothesis is that the ostensive approach (reflected in both POB and COB) is generally more acceptable or satisfying to the user. It can be further distinguished in two sub-hypotheses: 1) Query adaptation along with an ostensive interface provides a better environment for CBIR, and 2) providing an explicit control on the ostensive system results in better user satisfaction.

5.3. Participants

Since we wanted to test the system as close to real-life usage scenario, our sample user population consisted of 18 post-graduate design students. Responses to a pre-search questionnaire indicated that our participants could be assumed to have a good understanding of the design task we were to set them, but a more limited knowledge or experience of the search process. We could also safely assume that they had no prior knowledge of the experimental systems.

We met each participant at a time, each on a separate occasion for which the procedure was as follows:

- an introductory orientation session
- a pre-search questionnaire
- for each of the three systems in turn:
  - a training session on the system
  - a hand-out of written instructions for the task
  - a search session in which the user interacted with the system in pursuit of the task
  - a post-search questionnaire
- a final questionnaire

6. RESULTS ANALYSIS

6.1. Pre-search Questionnaire

Through this questionnaire, information about the participants' experience with computers and familiarity with using photographs was obtained. It revealed that all of the users worked extensively with images, and that they were often required to retrieve images from large collections.

6.2. Post-search Questionnaire

After each search task on one of the systems given a particular task, the users were asked to complete a questionnaire about the task they were given, the search they performed, the system they used, etc.

6.2.1. Semantic Differentials

Each respondent was asked to describe various aspects of their experience of using each system, by scoring it on the same set of 28 7-point semantic differentials. The differentials focused on five different aspects:

- three of these differentials focused on the *task* that had been set (part 1) (clear, simple, familiar);
- six focused on the *search process* that the respondent had just carried out (part 2) (relaxing, interesting, restful, easy, simple, pleasant);
- five focused on the set of images *retrieved* (part 3) (relevant, important, useful, appropriate, complete);

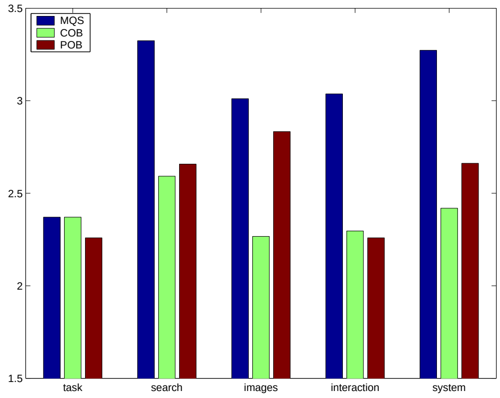


Fig. 3. Semantic differential means per part

- three focused on the user's perception of the *interaction* with the system (part 4) (in control, comfortable, confident); and
- eleven focused on the *system* itself (part 5) (efficient, satisfying, reliable, flexible, useful, easy, novel, fast, simple, stimulating, effective).

The result was a set of 1512 scores on a scale of 1 to 7: 18 respondents scoring each of 3 systems on each 28 differentials. Given the nature of the data, we had to use the non-parametric form of analysis of variance—the Friedman test—to analyse the results.

Overall, the Ostensive Browsers outperformed MQS, and usually COB's scores were lower (better) than the scores for its pure counterpart. The means of all differentials for each part are depicted in figure 3, which shows the trend that MQS scores are poorer than the scores for the other two systems, supporting our initial claim that query adaptation along with an ostensive interface provided a better environment for CBIR. The graph also shows quite clearly that POB's scores are comparable with COB, apart from the scores for part 3. This part focused on the retrieved images, thus backing up our second sub-hypothesis, namely that providing an explicit control on the ostensive system resulted in better satisfaction on the task completion. The results for the subset of differentials, which showed a significant level at $p <= 0.05$, are given in table 1. The table shows the differential, along with the mean score for each system, and the resulting p-value (after adjustment for ties). Dunn's multiple comparison post test was performed for each of the differentials to determine which systems' scores show a significant difference. The most significant results are found when comparing the differentials for the system part. Most notable is the variance in judging the systems' usefulness, and it should be pointed out that the advantage of the POB as being the simplest tool to use is reflected in the results, as well.

There were no significant differences for part 1 (concerning the tasks), neither across the systems nor across the tasks, which shows that the tasks were well-balanced and are believed not to have confounded the results significantly.

6.2.2. Likert Scales

Each user was asked to indicate, by making a selection from a 7-point Likert scale, the degree to which they agreed or disagreed with each of seven statements about various aspects of the search process and their interaction with the system.

There were four statements concerning the user's information need, which were phrased in such a way that responses would indicate the extent to which:

1. the user's initial information need was well-defined ("I had an idea of the kind of images that would satisfy my requirement before starting the search");
2. the user was able to find images representative or coextensive with that need ("The retrieved images match my initial idea very closely")
3. the user's information need changed in the course of the search ("I frequently changed my mind on the images that I was looking for"); and
4. the change of their need was due to the facilities offered by the system ("working through the image browser gave me alternate ideas").

The rest of the statements tried to capture the user's satisfaction with the search process and the system. They were phrased in such a way that responses would indicate the extent to which:

5. the user was satisfied with the outcome of the search ("I am very happy with the images I chose");
6. the user was satisfied with the level of *recall* attained ("I believe that I have seen all the possible images that satisfy my requirement"); and
7. the user was satisfied with the overall outcome of their interaction with the system ("I believe I have succeeded in my performance of the design task").

Like before, each user was asked to respond to these statements three times (after each task they carried out on the different systems). The result was a set of 378 scores on a scale of 1 to 7 (with 1 representing the response "I agree completely" and 7 "I disagree completely"): 18 respondents scoring each of three systems with respect to each of the seven statements. The mean results are shown in table 2.

Furthermore, since an evaluation based on the retrieved images *after* the search had been completed is hindered by subjective bias [1], the participants were invited to draw

Differential	MQS	POB	COB	p-value	Dunn's post test
Part 2:					
restful	3.9	3.1	2.8	0.008	MQS vs. COB < 0.05
pleasant	3.4	2.6	2.2	0.050	-
Part 4:					
comfortable	3.2	2.2	2.2	0.014	-
Part 5:					
flexible	3.7	3.4	2.4	0.007	MQS vs. COB < 0.05 POB vs. COB < 0.05
useful	3.4	2.6	1.9	0.001	MQS vs. COB < 0.01
novel	3.3	2.4	2.0	0.010	MQS vs. COB < 0.05
simple	2.9	2.2	2.9	0.030	-
stimulating	3.3	2.6	2.1	0.003	MQS vs. COB < 0.05
effective	3.2	2.4	2.1	0.007	MQS vs. COB < 0.05

Tab. 1. Means and significance test results (value range 1-7, lower=better)

statement	MQS	POB	COB
1	1.8	1.4	2.2
2	3.2	3.0	3.2
3	4.2	4.0	3.4
4	3.3	3.0	2.4
5	2.8	2.7	2.3
6	4.1	3.0	2.9
7	2.9	2.4	2.3

Tab. 2. The mean scores for each statement

sketches of the kind of images they had in mind before starting the search. This ensured that there was a point of reference for them to judge whether the retrieved images matched their initial sketches.

Information Need: The scores for the respondents' reactions to the first two statements did not vary very much across the systems. When they were asked about their initial information need, most of the user's stated that they had a rather clear idea of the images they were looking for (stmt. 1). Similarly, the responses for the second statement whether the retrieved images matched their initial information need, were uniform across the systems (stmt. 2). However, the next two statements indicate differences in the systems. The responses for COB showed that users were more inclined to change the initial information need (stmt. 3), whereas MQS and POB responses did not tend either way. When asked whether they thought the system gave them alternate ideas (stmt. 4), COB scored significantly better. The significance of the difference is reflected in the values of the Friedman test statistics calculated in order to test the experimental hypothesis that the scores for COB are better (lower) than for MQS. The value of the Friedman statistic was found to be significant at a level of $p < 0.05$ (p=0.024 after adjustment for ties).

Analysing the users' comments about why they thought the images matched their initial idea (stmt. 2) and why they changed their idea (stmt. 3) sheds more light on the above results. To summarise, it emerges that the COB supports an explorative search, which is greatly appreciated by the users. In MQS however, many people at some point faced the problem that they were unable to retrieve any more images (usually when they exhausted keywords). They often had the feeling the images they were looking for were not in the database, and they were puzzled because they could not tell whether the images were indeed not there or whether they could not formulate the query properly. The majority of people who changed their mind on the initial images interpreted that in a negative way as a result of not being able to find the right ones. One person's comment reflects this mood: *"My expectations have been adapted to the available images. This, however, is not how a designer wants to think, he doesn't want limitations to influence decisions.*"

6.3. Final questionnaire

After having completed all three tasks, the participants ranked the three systems in order of preference with respect to (i) the one that *helped* more in the execution of their task, and (ii) the one they *liked* best. Both questions resulted in a very similar ranking. 15 out of the 18 participants ranked COB more highly than the other systems, and 12 placed both ostensive interfaces as their top two. The mean of the ranks were: MQS 2.5, POB 1.9, and COB 1.6. Again, in order to test the experimental hypothesis that the sets of 18 post-search scores for each system type were sampled from different populations, the Friedman statistic was calculated, which was found to be significant at a level of $p = 0.017$ (for both questions (i) and (ii)). Dunn's post test showed that the significant difference was between MQS and COB (with $p < 0.05$). Our conclusion, therefore, was that people liked COB significantly better than MQS, and found it significantly more useful for the task we set them.

Respondents who ranked MQS highest appreciated the system's accuracy and being able to control the search—e.g. *"fastest of the 3 systems in finding specific images"*. On the other hand, the responses of people who preferred one of the other systems showed the complexity of formulating a query in MQS: *"quite complex"*, *"have to input too often"*, *"confusing to control"*. Some people also found MQS *"too restrictive"*.

Those respondents who preferred either of the ostensive browsing approaches valued the fact that they were intuitive (*"ease is good"*, *"more intuitive to first time users"*) and the graphical display of the images (*"easily understandable 'line of choices'"*, *"ability to compare images on screen + backtracking"*). The advantages of the pure version (POB) were stated as its simplicity and *"pleasure to use"*: *"more fun, and still effective and fast"*, *"simple + fast"*, and *"very fluid movement—just the images"*. POB's drawbacks were concentrated on the missing ability to control the search: *"does not allow enough input/editing [...] during image browsing"* and *"doesn't give enough info"*. The additional control options, however, were also criticised by some users in the COB: *"too much choice of search options"*. Apart from this, most responses about COB were entirely positive. People liked its adaptability and versatility: *"it has the option of control, be it whether I use it or not"*, *"most options, best display of information"*, *"gave lots of alternatives by altering search parameters"*. In addition, the responses showed the effectiveness of the system: *"it is most efficient to use and get the desired results"*, *"search seemed more consistent"*, and *"felt more extensive"*.

A CBIR system needs to be flexible enough to accommodate different types of users and search requirements. We believe our initial analysis shows the success towards a consistent, effective and versatile approach. Again, a user's comment summarises this ability: *"I liked the flexibility when I needed and the automatic selection when I didn't"*.

7. CONCLUSION

We developed and described an adaptive approach towards CBIR. We used both text and colour features and combined them using the Dempster-Shafer theory of evidence combination. We furthermore presented a user-centred, task-oriented evaluation of the systems, which demonstrated the value of our technique by comparing it to a traditional CBIR interface.

In recognition of the fact, that information seeking is an inherently interactive activity, the system should provide for an intuitive and interactive interface. For this reason, we believe more emphasis has to be put on the human computer interaction aspects to design a supportive system, rather than attempting a 100% accuracy in the retrieval.

8. ACKNOWLEDGEMENTS

The authors wish to thank the searchers who participated in the study. This research was supported by the EPSRC (U.K.) research grant GR/R74642/01.

9. REFERENCES

[1] J. A. Black, Jr., G. Fahmy, and S. Panchanathan. A method for evaluating the performance of content-based image retrieval systems based on subjectively determined similarity between images. In *CIVR 2002, LNCS 2383*, pp 356–366, 2002.

[2] I. Campbell. Interactive evaluation of the Ostensive Model, using a new test-collection of images with multiple relevance assessments. *Journal of Information Retrieval*, 2(1):89–114, 2000.

[3] I. Campbell. *The ostensive model of developing information needs*. PhD thesis, University of Glasgow, 2000.

[4] I. Campbell and C. J. van Rijsbergen. The ostensive model of developing information needs. In *Proc. of CoLIS-96*, pp 251–268, 1996.

[5] M. Chalmers, K. Rodden, and D. Brodbeck. The order of things: Activity-centred information access. *Computer Networks and ISDN Systems*, 30(1–7):359–367, 1998.

[6] J. M. Jose. *An Integrated Approach for Multimedia Information Retrieval*. PhD thesis, The Robert Gordon University, Aberdeen, 1998.

[7] J. M. Jose, J. Furner, and D. J. Harper. Spatial querying for image retrieval: a user-oriented evaluation. In *Proc. of the 21st annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp 232–240, 1998.

[8] J. Peng, B. Bhanu, and S. Qing. Probabilistic feature relevance learning for content-based image retrieval. *Computer Vision and Image Understanding*, 75(1–2):150–164, 1999.

[9] K. Porkaew, K. Chakrabarti, and S. Mehrotra. Query refinement for multimedia similarity retrieval in MARS. In *Proc. of the 7th ACM Int. Conf. on Multimedia*, pp 235–238, 1999.

[10] J. J. Rocchio. Relevance feedback in information retrieval. In G. Salton, editor, *The SMART retrieval system: experiments in automatic document processing*, pp 313–323. Prentice-Hall, Englewood Cliffs, US, 1971.

[11] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Tran on Circuits and Systems for Video Technology*, 8(5):644–655, 1998.

[12] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Tokio, 1983.

[13] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[14] A. H. M. ter Hofstede, H. A. Proper, and T. van der Weide. Query formulation as an information retrieval problem. *The Computer Journal*, 39(4):255–274, 1996.

[15] X. S. Zhou and T. Huang. Relevance feedback in image retrieval: A comprehensive review. *ACM Multimedia Systems Journal, Special Issue on CBIR*, 8(6):536–544, 2003.